

A matematikai statisztikáról

Hogyan kerülhetek az ünneplők közé és miért ajánlom munkámat kedves barátomnak, Szentpéteri Istvánnak? Elsősorban a magam okulására, mert ő volt az, aki az elsők között hívta fel figyelmemet a kutatásra, a tudományos tevékenység jelentőségére. Annak ellenére, hogy én matematikus voltam, ő meg jogász, el tudta velem hitetni, hogy fontos az elmélyült vizsgálódás minden területen. Abban, hogy hittem neki, bizonyára szerepet játszott az a tény, hogy egyazon cserkészcsapat tagjai voltunk – még gimnazista koromban – és ő mint jó vezető, sok mindenben segített, figyelte fejlődésemet. A matematikai szakmai támogatást Szegeden ugyancsak volt makói diákoktól, a nagyon tehetséges és kitűnő pedagógusoktól, Pintér Lajostól és Szendrei Jánostól kaptam. Ezúton is köszönöm támogatásukat, figyelmüket, hiszen nélkülük aligha váltam volna azzá, amiről akkor nem is álmodtam. Kinyitották a szemem egy új világra. Nagyon szeretném hinni és remélni, hogy nekem is sikerült néhány szemet "kinyitni".

Nem meggyőzni – vagy pláne kioktatni – szeretném a nem matematikusokat, inkább fordítva, a matematikusok figyelmét kívánom egy kicsit irányítani a humán feladatokra. Néhány olyan megfigyelésem szeretném megosztani az olvasókkal amelyeket magam érdekesnek tartottam, vagy néha ilyen-olyan okból tanulmányoztam. Megemlítve, hogy a törvényszéki és bűnüldözési munkában felhasznált statisztikai eszközökre még a hatvanas években Kertész Imre hívta fel a figyelmemet, tőle sokat tanultam és tanácsára az irodalmat – elsősorban angol nyelvűt – is tanulmányoztam, reményeim szerint haszonnal. Akkor jöttek divatba a számítógépek és a számítógépes feldolgozások. Az ujjenyomat-értékelés és az írásszakértés voltak a vizsgálódások tárgyai. Sikerült máig érvényes megállapításokat tennünk. Az ujjenyomat-vizsgálat statisztikailag statikus feladat, és a legnagyobb probléma a szakértők felismerő tevékenységének algoritmizálása. Az írásszakértésben a mozgás folyamata az elsődleges, aminek gépi felismerése máig megoldatlan, a sztochasztikus folyamatok elméletét, az idősorok elemzését igényli. A megfelelő modell és algoritmus elkészítését a gépbe vitel és feldolgozás elé helyeztük.

Az akkor megfogalmazott feladatok megoldása mára igen magas szintet ért el, ezekről nem kívánok írni, csak nagy meglepéssel konstatálni, hogy a matematikai statisztikai megfontolások máig szükségesek és a gépi feldolgozások alapjai. Az első ilyen általános elv, amit a megfontolások alapjának tartok: a statisztikai egyensúly vizsgálata és értelmezése. Ebben a környezetben, mármint humán beállítottságú emberek, jogászok, szociológusok között sem tűnik nehéznek vagy reménytelennek matematikáról beszélni. Hiszen a törvényszéki gyakorlatban, a helyzetek megítélésében, de a mindennapi életben is használjuk a „majdnem biztos”, a „szinte bizonyos”, a „lehetetlen” vagy éppen „a botnak is két vége van” kifejezéseket. Legfeljebb nem gondolunk arra, hogy ezek a megállapítások számszerűsíthetők is. Sőt, matematikai model-

lek alkothatók, amelyek segítségével sokminden vizsgálható, előre jelezhető. Az említett ítéletek mindegyikének valószínűségi, vagy más szóval, matematikai statisztikai hátere van, ami egyben azt is jelenti, hogy a mérések elvégzése nemcsak bonyolult hanem költséges és igen nagy körültekintést igényel. Többször használok a "mérés" kifejezést, amit szintén meg kell magyarázni, mert statisztikai fogalom: mérés alatt értjük a statisztikai adatgyűjtést és feldolgozást. Ha még ehhez számítógépes feldolgozás is szükséges, amit a gondos feladatleírás igényel, máris a költségeknél tartunk, hisz a költségek megsokszorozódhatnak. A jogot, társadalmi viszonyokat érintő feladatok pontos megfogalmazása – úgy érzem – szükséges és fontos feladata a tudománynak és a gyakorlatnak is. Úgy vélem, legjobb, ha az alapvető fogalmakat példákon keresztül mutatom be, világítom meg.

A bizonyosságról

Annak lehetőségét, hogy vétlen embert ítéljenek el, logikailag nem lehet kizárni, bármilyen legyen is a jogrendszer, kivéve azt az extrém esetet, hogy senkit nem ítélnek el. Ebben a különleges esetben minden bűnöst futni hagynak, bármilyen büntényről legyen is szó. Elképzelni is borzasztó, mihez vezetne egy ilyen rendszer. Pedig itt egy olyan alapvető problémáról van szó, amelyet ugyancsak helyes lenne számszerűsíteni, a tényleges helyzettel összevetni. Az a kívánság ugyanis, hogy csak a 100%-os, azaz a teljesen biztos esetben lehessen ítélkezni, vagy akár csak véleményt mondani, lehetetlen követelmény. Valamilyen egyensúlyban kell tartani a kétfajta hiba előfordulási valószínűségét: a vétlen elítélése – elsőfajú hiba – amelynek a valószínűségét α -val jelöljük, a bűnös futni hagyása – ez a másodfajú hiba – amelynek a valószínűségét β -val jelöljük, előfordulásának statisztikai vizsgálatára van szükség, ezek alapján lehet gyakorlati tanácsokat adni, vagy éppen döntéseket hozni. A különböző bűnelkövetések esetén ezek a valószínűségek más és más gyakorlatot követnek. Lopás esetén a bizonyítékok felderítése és bemutatása, a tettes kilétének kiderítése jóval kevesebb utánjárást mutat a gyakorlatban, mint pl. a gyilkosságok felderítése. Ez a gyakorlat két módon mutatkozik meg a lopásoknál:

(a) sokkal nagyobb számú a vétkesek futni hagyása, azaz többen maradnak büntetlenek a bizonyítások elégtelensége miatt, mint a vétlenek elítélése. Erre mondjuk matematikus nyelven: nagy a másodfajú hibavalószínűség és igen kicsiny az elsőfajú hibavalószínűség.

(b) Más szempontból vizsgálva ugyanezt a jelenséget: a lopások száma jóval nagyobb, mint a felderített esetek száma. Ebben az utóbbi esetben a felderítés valószínűségének időbeli változása az, ami fontos és tanulmányozandó. Ekkor lehet dönteni arról, hogy a társadalomra nézve mennyire veszélyes jelenségről van szó. Lehet trendet vizsgálni, idősor elemzést végezni és az okokat is feltárni.

Más a helyzet a gyilkosságokkal, ebben az esetben mind az elsőfajú hibavalószínűség (ártatlan elítélése), mind a másodfajú (a gyilkos el nem ítélése) hibavalószínűség kicsiny kell, hogy legyen (mindkettő nem tehető nullává!). Ez pedig a bizonyítékok alapos, mindenre kiterjedő összegyűjtését igényli. A költségeket lehet mérni, de azokat nem lehet a bűn nagyságával hasonlítani. Más mérőszámra van szükség. Mint mérési lehetőség vethető fel (a halálbüntetés eltörlése miatt számolni kell a gyilkosok szabadu-

lásával), hogy ha pl. a szabaduló gyilkosok 0.5 valószínűséggel újból gyilkolnak (ez a valószínűség lehet, hogy kisebb) és ha évente mintegy 50 gyilkos szabadulásával kell számolni, akkor is az ebből adódó vétklen áldozatok száma többszöröse lesz a vétklen elítéltekének. Az emberi jogok említésekor az ne csak a gyilkosokra vonatkozzon, hanem a társadalom ilyen károsultjaira is. A felderített esetek száma ebben az esetben is összemérendő az összes esettel, nyilván sokkal több munkát és energiát igényel a 80–90%-os felderítési arány, mint a lopásoké, különösen 1 esetre vonatkozóan. És ez így természetes is, még ha a lopáskárosultaknak ez nem esik jól. Mindent felderíteni még ebben az esetben is, mint elvárás lehetetlen, bármilyen anyagi lehetőségekről legyen is szó. Ez nem jelenti azt, hogy a megelőzésről nem kellene jobban gondoskodni, dehát ez a gyilkos felderítésénél is bonyolultabb feladat. Ha a gyilkosságok esetén a helyzet évente romlik, azaz jelentősen megnő a számuk, akkor alapos vizsgálatra van szükség, hogy az okokat fel lehessen tárni. Megvizsgálandó például, hogy annak következménye-e a változás, hogy a büntetés mértéke nem megfelelő, vagy a társadalmi mozgásokban történt változás vagy a társadalom összetétele változott, esetleg egyéb ok miatt történt lényeges átalakulás. Ez esetben mivel mérjük a szigorítás mértékét? Még az olyan "dogmaszerű" érvelés is nehezen tartható, hogy a halálbüntetést azért kell eltörölni, mert mindenkinek joga van az élethez. Hiszen mi lesz azoknak a jogával, akik teljesen vétklenek, és gyilkosságok áldozatai, éppen a számszerű növekedés miatt azok, mint azt fentebb már mutattam? Kevés lenne azzal megelégedni, hogy csak azon feltétel melletti növekedést kell vizsgálni, hogy a halálbüntetést eltörölték. És jogos a kérdés – azt is hogyan?

Remélhető, hogy a teljes bizonyosság kérdése statisztikai értelmezése utat tör magának a mindennapi gyakorlatban és nemcsak elvi szinten. Különösen érdekes lehet a már említett eszközökön kívül az olyanok alkalmazása, amelyre eddig nem volt egyáltalán lehetőség, mint pl. a többdimenziós analízis felhasználására amint azt a társadalomtudományokban gyakran teszik (szociológia, közvéleménykutatás).

Új folyamatok megjelenése

Egyszerűen megfogalmazható kérdés, de nem kevésbé fontos, hogy milyen következményei vannak a határokon való ki- és belépés könnyebbé tételének a bűnüldözésben, a jogalkotásban és jogalkalmazásban? Kezdjük a legegyszerűbbel, a költségekkel. Ha nő a bűnelkövetések száma, mint ahogyan nő, az ugyanolyan szinten tartás megkövetelése a felderítés, a bizonyítás költségét növelni fogja. Azonos költségvetési ráfordítások esetén katasztrofálisan növekedni fog a felderítetlen esetek száma. Ez a szám megint a másodfajú hibavalószínűséggel mérhető, tehát a meg nem büntetett elkövetők aránya fog növekedni – kérdés, meddig engedhető meg? Vagy mennyivel kell a pénzügyi terheket növelni egy tűrhető szint eléréséhez? Miért is nő a bűncselekmények száma? Két okot gyorsan meg lehet említeni: a beérkező külföldiek között sok az alvilági elem, a hazai alvilág külföldre menekülhet. A bizonyítási eljárásokban minőségi változást jelent az a tény, hogy a korábbi tíz és félmillió hazai lakosság helyett egy nagyságrenddel nagyobb lehetséges "tömegeg" kell számolni. És ezzel együtt nő a lehetséges elkövetők száma is. Az utóbbi esetben nemcsak a statikus valószínűségeket kell meghatározni, hanem a "folyamatok" időbeni lefolyását, azaz trendeket – azok megbízhatóságát – szezonális komponenseket, és a ráakódó vétkleneket is.

A valószínűségekről

A valószínűségek meghatározásával, azok jelentőségével és felhasználhatóságának korlátaival igen sokat foglalkozik a törvénykezés szakmai és tudományos irodalma. A leggyakrabban előforduló bizonyító erejű eseményekkel kapcsolatban használatos valószínűségek pl. a törvényszéki gyakorlatban a következők:

- vércsoportokra vonatkozó,
- DNA csoportok (deoxyribonucleinsav),
- anyagmaradvány, mint pl. szövet, üveg, cserép,
- hajszál, sperma,
- balkezesség, színvakság, testi fogyatékoság.

Egy esemény valószínűségét $P(\cdot)$ -val jelöljük, az eseményeket latin nagybetűkkel, gyakran index felhasználásával, például a vércsoportok esetén:

V_1	A-típus	$P(V_1)=0.417,$
V_2	B-típus	$P(V_2)=0.104,$
V_3	AB-típus	$P(V_3)=0.039,$
V_4	0-típus	$P(V_4)=0.441.$

Ezeket a valószínűségeket statisztikai mérésekkel határozták meg. Megemlítve, hogy ilyen és hasonló táblázatokat minden országban vezetnek, pl. a kínai adatok a fentiektől eltérnek. A fenti valószínűségek alapján pl. hazánkban kb. 390 ezer az AB-típusú vércsoportba tartozók száma. Annak valószínűsége, hogy egy házaspár ehhez a vércsoport-hoz tartozzon $0.039 \times 0.039 = 0.0015$, (tízezer párból 15). Általában is ha csak azt kérdezzük, hogy mennyi annak valószínűsége, hogy két véletlenül kiválasztott egyén azonos vércsoportba tartozzon, a fenti valószínűségek négyzetei összege: 0.3807. Ahhoz, hogy megfelelő döntést lehessen hozni egy gyanúsított hovatartozására vonatkozóan, több, független tulajdonság valószínűségét lehet felhasználni. Esetünkben, ha mérni lehet a vércukor tartalmat a helyszínen és a gyanúsítottnál, továbbá a vörös vérszámát és esetleg más tulajdonságokat is a gyanúsítottnál és a helyszínen találtnál és az egybeesés valószínűsége egy ezrednyi lesz, úgy ez igen erős bizonyító erőt jelent (az elsőfajú hiba- valószínűséget kapjuk).

Más szóval ezt úgy fejezzük ki, hogy az adott X, Y, Z feltételek mellett a gyanúsított bűnösségének valószínűsége $P(B|X, Y, Z)$. Egy A esemény ellentétjét

(kiegészítőjét) A^C jelöli $P(A), P(A^C) = 1 - P(A)$. Szokás az összehasonlító bizonyító erőről (angolul odds) is beszélni, ami alatt a két valószínűség hányadosa értendő, pl. az AB vércsoport esetén: $0.961/0.039=24.6$. Ennek kapcsán összetettebb események valószínűségét is meg lehet határozni. Mérések alapján, vagy adott feltételek mellett az első és másodfajú hiba valószínűsége meghatározható. Használva a feltételes valószínűség fogalmát és a Bayes-féle tételt, összetett események valószínűségei is meghatározhatók. Tekintsük a következő egyszerű problémát. Két anyagmaradvány vizsgálatakor több mérést végeznek el, pl. üvegtörmelék egy betört ablaknál és egy gyanúsított ruháján. Származhat-e a két törmelék ugyanabból az ablakból? Továbbá hogyan

változik a bűnösségre vonatkozó odds értéke és az α, β hibavalószínűség? A mérések pl. az üvegtörmelék visszaverő képességét (pontosabban a törésmutatót) mérik. A kis darabok mérési hibával adják a visszaverődési értékeket. Tegyük fel, hogy mérési eredmények Gauss eloszlásúak θ_1 illetve θ_2 középértékkel és ugyanazon σ szórással. Ha a helyszínen m mérést végzünk, továbbá n legyen a gyanúsítottnál talált anyag méréseinek száma. Azonosság esetén $\theta_1 = \theta_2$, egyébként $\theta_1 \neq \theta_2$. Jelölje X a helyszíni eredményeket, Y a gyanúsítottnál mértéket. Abban az esetben, ha a tükrözési mutatók a szokásosak, nincs semmi különlegesség bennük, nem sokat jelent $\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i$ és

$\bar{Y} = \frac{1}{n} \sum_{ji=1}^n X_{ji}$ közelsége. Ha viszont a mutatók ritka jelenségre utalnak, az \bar{X} és

\bar{Y} közelsége esetén a két mérés ugyanarra a sokaságra utal, azaz a két törmelék egy ablakból származik, nagy megbízhatósággal. A törésmutató valódi értékét is Gauss-eloszlásúnak szokás tekinteni μ középpel és τ szórással. A Gauss-ság feltétel nem lényeges, a számolások elvégzését könnyíti. Most ezt matematikai jelölésekkel is elvégezzük.

Legyen a helyszíni mérések jelölése X_1, X_2, \dots, X_m , a gyanúsított mérési eredményei pedig legyenek Y_1, Y_2, \dots, Y_m , azaz azonos számú megfigyelésünk van,

bár ez nem lényeges megszorítás. \bar{X}, \bar{Y} jelöljék az átlagokat, $\left(\tau^2 + \frac{\sigma^2}{m} \right)^{1/2}$ a

szórásukat, akkor feltéve, hogy a bűnösség $B=I$, (I – a törmelék azonos forrású), akkor odds-a az addigi bizonyítékok (E) alapján és X, Y alapján a következő lesz:

$$\frac{P(B|X,Y,E)}{P(B^c|X,Y,E)} = \frac{p(X,Y|B,E)}{p(X,Y|B^c,E)} \frac{P(B|E)}{P(B^c|E)} = \frac{p(X,Y|I)}{p(X,Y|I^c)} \frac{P(B|E)}{P(B^c|E)},$$

feltéve, hogy $E, (X, Y)$ függetlenek. Tehát az eredeti bizonyító erő szorozódik az azonosság illetve a különbözőség sűrűségfüggvények hányadosával. A sűrűségek hányadosa pedig a Bayes-feltétel mellett a törésmutató eloszlásától függ, ha a szórás

$\tau \gg \frac{\sigma}{\sqrt{m}}$ lényegesen nagyobb, akkor a számítások alapján a következő képlet alap-

ján lehet számolni:

$$\bar{\sigma}^2 = \tau^2 + \frac{\sigma^2}{m}, a^2 = \frac{2}{m}, \bar{Z} = \frac{1}{2}(\bar{X} + \bar{Y}),$$

$$\frac{p(X, Y|I)}{p(X, Y|I^C)} = \frac{\tau}{a\sigma} \exp\left(-\frac{(\bar{X} - \bar{Y})^2}{2a^2\bar{\sigma}^2} + \frac{(\bar{Z} - \mu)^2}{2\tau^2}\right).$$

Ha itt $\frac{\sqrt{m\tau}}{\sigma} = 100$, a Bayes féle tényező (szorzó) több száz is lehet. Ennek alapján számolva az első és másodfajú hibavalószínűséget az jelentősen kisebb lesz, mint az X és Y mérések nélkül.

Az odds, összehasonlító bizonyító erő növelésének másik lehetséges módja a gyanúsítottak körének szűkítése. Ez elérhető pl. megfelelő szemtanúk felkutatásával. Az ezzel kapcsolatos számítások ugyancsak elemiek, de semmiképpen nem triviálisak.

A technikai eszközök fejlettsége, a számítógépek új lehetőségeket, új utakat biztosítanak a fenti típusú feladatok megoldásában, de csak akkor, ha a választott matematikai modell valóban adekvát.

Még egy jelentős változásra szeretném felhívni a figyelmet, amelynek mérése, számszerűsítése kívánatos lenne. Ez a gazdaságosság számítása és alátámasztása statisztikai adatokkal. Itt a változó körülmények, a lényegesen változó új bűnözési formák éppúgy figyelembe veendők, mint a változó oktatás és egészségbiztosítás. A múltban elképzelhetetlen volt az olyan mértékű olajcsalás kialakulása, mint ami néhány hónap alatt az átmeneti időszakban kialakult, ez egy szignifikáns trend megjelenése volt, inspirálója a gyors meggazdagodás, oka – többek között – a benzin árába beépített túl magas adó. Itt nem használhatók az idősor-elemzés eszközei. Kezelése, leleplezése még ma is gyerekcipőben jár. Az autólopások száma olyan méreteket öltött, hogy a biztosítók nem kötnek CASCO biztosítást. Hogyan lehetne a megelőzést úgy megszervezni, hogy a biztonság növekedjen, és a kiadások csökkenjenek?

MÁTYÁS ARATÓ

ON MATHEMATICAL STATISTICS

(Summary)

In this paper we discuss some statistical problems of law, and forensic science. We provide the solutions for many problems by the help of the two probabilities of error. The connection, relationship of the problems to significance tests is also shown. We consider how the results might be used in presenting evidence in a court of law or in everyday life.